DOCUMENT RESUME

ED 173 440 TM 009 618

AUTHOR Loveridge, Robert L.; Carapella, Ruth
TITLE The Application of USOE Models A, B, & C to a Title I Early Childhood Program.
PUB DATE Apr 79
NOTE 17p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, April 8-12, 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; *Achievement Gains; Admission Criteria; *Compensatory Education Programs; *Data Analysis; Early Childhood Education; Evaluation Criteria; *Evaluation Methods; Kindergarten Children; Models; Preschool Tests; Pretests; *Program Evaluation; *Research Design; Test Selection; Underachievers
IDENTIFIERS Elementary Secondary Education Act Title I; *Kindergarten Extended Day Program

ABSTRACT

Kindergarten pupils enrolled in Elementary and Secondary Education Act Title I schools were selected for the Kindergarten Extended Day (KED) Program if they showed marked underachievement on the Cooperative Preschool Inventory (CPI). KED students attended kindergarten for an additional half day. Since previous evaluations of KED may have had problems in their research design, this evaluation used and compared six of the Office of Education Title I evaluation models. The following comparisons were made, using transformed data from the CPI, Basic School Skills Inventory (BSSI), and the Comprehensive Tests of Basic Skills (CTBS)--including Visual and Auditory Discrimination and Mathematics subtests: (1) Model A comparison of normal curve equivalent gains on BSSI for KED participants; (2) Model C, using CTBS subtests as criterion variables and CPI as the selection instrument; (3) modified Model B, using BSSI and CTBS posttest scores and CPI selection scores; (4) alternative Model B, using CPI selection scores as a covariate; (5) Model C evaluations using KED participants as in (3); and (6) Model A evaluation as in (3). Results indicated that substantially different results were produced from one model to another. Six recommendations were made regarding program evaluation models. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED173440

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

THE APPLICATION OF USOE MODELS A, B, & C TO A TITLE I EARLY CHILDHOOD PROGRAM

Robert L. Loveridge, Ph.D.
Ruth Carapella, Ph.D.

St. Louis Public Schools
St. Louis, Missouri

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Robert L. Loveridge

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

TM009 618

Presented at the 1979 Annual Meeting of the American Education Research Association, San Francisco, California, April 8-12, 1979.

The Kindergarten Extended Day (KED) program was implemented as a Title I component in the Fall of 1974. Kindergarten pupils enrolled in Title I schools which offered KED were selected for participation on the basis of evidenced academic underachievement, as determined by the Cooperative Preschool Inventory (CPI). Those children who scored at or below the 50th percentile of children aged 5 years 6 months to 6 years 5 months were, with parental permission, identified for the program. Pupils with the lowest scores in each school, were to be selected first.

Participating pupils attend kindergarten for an additional half day. Instruction supplements regular kindergarten instruction in the areas of language arts, mathematics, and social development.

Evaluation of the effects of the KED program has been obscured for various reasons since the program's inception, and while this report by no means represents the perfect evaluation, it has been able to produce more answers than previous evaluations. This was possible mainly because during the 1977-78 school year (1) selection and pretest were different instruments, (2) class size limitations made it possible to identify a control group which had not received KED services, and (3) the NCE metric (while not without problems of its own) was utilized to evaluate achievement gains.

Preliminary analysis, based upon raw score data of the following instruments - the Cooperative Preschool Inventory (CPI), the Basic School Skills Inventory (BSSI), and two subtests, Visual & Auditory Discrimination (VAD) and Mathematics (Math), of the Comprehensive Tests of Basic Skills (CTBS), indicated that:

A. those pupils enrolled in KED had large pre to posttest gains as measured by the BSSI.

B. for a matched group of students, selected post hoc, as a

comparison group for KED students, KED students had significantly higher raw scores on the BSSI.

C. also on the same comparison groups, when differences in selection scores were controlled, KED pupils evidenced higher scores on two subtests of the CTBS, Math and VAD.

While points 2 and 3 above suggested that the KED program was more effective than no treatment, analysis of raw scores alone could obscure the actual program effects (e.g., maturation effects are not controlled in an analysis of raw scores). Raw scores were utilized in the preliminary analysis because the publisher of the BSSI did not provide norms suitable for Title I evaluation (i.e., percentiles, NCE's).

The purpose of this paper was to apply the various Title I models to the data to determine if differences resulted and if results were comparable across models. By special arrangement with the BSSI test developer we were able to secure the raw data used in standardizing the instrument for the four age groups of our study. Given the raw data we were able to compute percentiles and NCE scores for each of the four age groups of interest for the BSSI. Standard conversion tables were used to convert percentile scores to NCE's for both the CPI and the subtests of the CTBS.

The following comparisons were made using the transformed data:

1. A Model A evaluation which compared NCE gain on the BSSI for pupils enrolled in KED.
2. A Model C evaluation using the VAD and Math subtests of the CTBS as criterion variables and CPI as the selection instrument.

3. A modified B evaluation - using BSSI and CTBS posttest scores and CPI selection scores. Although the control group was not identified prior to treatment, because of class size limitations in participating schools, we were able to construct a post hoc comparison group of pupils with similar selection scores and with the same home school base.

4. An alternative to Model B on the students mentioned in 3 above wherein a covariate, CPI selection scores, was employed in the analysis.

5. A Model C evaluation on the KED students in 3 above.

6. A Model A evaluation on the KED students in 3 above.

RESULTS

1. The Model A evaluation results revealed an NCE gain of 31.20 for those pupils enrolled in the KED program and below the cutoff score on the selection instrument. This would imply that the KED program was tremendously effective in remediating evidenced academic underachievers since the gain was approximately equal to 1.5 standard deviations.

2. A Model C evaluation, using the entire population of KED pupils (1,372 pupils) was conducted. Results of the evaluation reported in Table 1 show mixed results for the KED program on the two CTBS subtests, VAD and Math. Positive treatment effects for KED pupils were found on both the VAD and Math subtests; however, using the 7 point NCE recommendation of the model developers, only the VAD results could be considered as educationally significant in favor of the treatment group.

TABLE 1

Predicted, Observed, and Differences in NCE's
on the Criterion Tests for KED Pupils

	Predicted Post	Observed Post	Difference
VAD*	32.64	40.04	7.40
Math†	41.30	43.40	2.10

* Slope = .47, control mean 49.06, pretest difference 34.93

† Slope = .34, control mean 53.18, pretest difference 34.93

3. Modified Model B evaluation results on the 507 students in the experimental study study varied according to the criterion variable employed. Table 2 shows that the ANOVA for the BSSI produced significant results in favor of the treatment group.

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE
(Post BSSI NCE Scores by Treatment Groups)

Source of Variation	SS	df	MS	F
Main Effects - Program	23225.78	1	23225.78	96.80*
Explained	23225.78	1	23225.78	96.80*
Error	106766.75	445	239.93	
Total	129992.56	446	291.46	

* Significant at .001 alpha level

Program Means

Treatment Group	(N = 272)	51.39
Control Group	(N = 175)	36.62

The treatment group achieved 14.77 NCE's above their control group counterparts. This was in spite of an initial difference in NCE scores of 13.29 in favor of the control group on the selection variable, CPI. Due to the manner in which the random samples were selected (i.e., in proportion to the number of students in each score group in their subpopulations) it was expected that there would be some difference on the selection score in favor of the control group. Although these differences existed when the ANOVA's were calculated for the two CTBS subtests (see Tables 3 and 4), there was no significant differences between treatment and control groups on the criterion variables.

TABLE 3

SUMMARY OF ANALYSIS OF VARIANCE
(CTBS Visual and Auditory Discrimination NCE Scores by Treatment Groups)

Source of Variation	SS	df	MS	F
Main Effects - Program	342.19	1	342.19	.94*
Explained	342.25	1	342.25	.94*
Error	178652.25	490	364.60	

* Not significant

Program Means

Treatment Group	(N = 268)	41.23
Control Group	(N = 224)	42.90

TABLE 4

SUMMARY OF ANALYSIS OF VARIANCE
(CTBS Math NCE Scores by Treatment Groups)

Source of Variation	SS	df	MS	F
Main Effects - Program	138.20	1	138.20	.35*
Explained	138.25	1	138.25	.35*
Error	190920.13	489	390.43	
Total	191058.38	490	389,915	

* Not significant

Program Means

Treatment Group	(N = 264)	46.45
Control Group	(N = 227)	47.51

4. A modified Model B with a Covariate was employed on each of the three criterion measures for the 507 experimental students, treatment and control. Since equality of slopes is assumed in covariance procedures, the slopes on each criterion variable with the selection test score were computed. The slopes were computed for those students above the selection cutoff score in the control group and for those students below the selection cutoff score in the treatment group. Table 5 shows these slopes.

TABLE 5

Slopes of Criterion Tests with Selection Test -
NCE Scores Based on Selection Test (CPI) - NCE Cutoff Score

Criterion Tests

	BSSI	CTBS	
		VAD	Math
Treatment Group	.39	.59	.41
Control Group	.43	.78	.49

With the possible exception of the VAD subtest the slopes for the two groups on the other variables were very similar. Results of the ANCOVA can be found in Tables 6, 7, and 8.

TABLE 6

SUMMARY OF ANALYSIS OF VARIANCE
(Post BSSI NCE Scores by Treatment Groups with CPI Scores as Covariate)

Source of Variation	SS	df	MS	F
Covariate - CPI	504.92	1	504.92	2.41*
Main Effects - Program	36584.68	1	36584.68	174.85†
Explained	37089.63	2	18544.81	88.63†
Error	92902.94	444	209.24	
Total	129992.56	446	291.46	

* Not significant

† Significant at .001 alpha level

Program Means

Treatment Group	(N = 272)	54.07
Control Group	(N = 175)	32.45

TABLE 7

SUMMARY OF ANALYSIS OF VARIANCE
(CTBS Visual and Auditory Discrimination NCE Scores
by Treatment Group with CPI Scores as Covariate)

Source of Variation	SS	df	MS	F
Covariate - CPI	14816.07	1	14816.07	44.85*
Main Effects - Program	2957.81	1	2957.81	8.95†
Explained	17773.94	2	8886.97	26.90*
Error	161205.81	488	330.34	
Total	178979.75	490	365.27	

* Significant at .001 alpha level

† Significant at .01 alpha level

Program Means

Treatment Group	(N = 267)	44.64
Control Group	(N = 224)	38.83

TABLE 8

SUMMARY OF ANALYSIS OF VARIANCE

(CTBS Math NCE Scores by Treatment Group with CPI Scores as Covariate)

Source of Variation	SS	df	MS	F
Covariate - CPI	9122.26	1	9122.26	24.71*
Main Effects - Program	2113.33	1	2113.33	5.73†
Explained	11235.63	2	5617.81	15.22*
Error	179770.13	487	369.14	
Total	191005.75	489	390.61	

* Significant at .001 alpha level

† Significant at .01 alpha level

Program Means

Treatment Group	(N = 263)	49.20
Control Group	(N = 227)	44.28

Results indicate that on the BSSI the adjusted scores shows a 21.62 NCE difference in favor of the treatment group. For this VAD subtest of the CTBS, there was a 5.81 NCE difference in favor of the treatment group. On the Math subtest of the CTBS there was a 4.92 difference in favor of the treatment group. While only the BSSI results would be considered educationally significant, the adjustments made on the criterion variables through the use of the covariate are different from the modified Model B by 6.59 NCE's on VAD and 5.98 NCE's on Math.

5. A Model C evaluation on the 507 pupils in the experimental study was also performed. This was done for two reasons: (1) it would permit the comparison of Models A, modified B, C, and modified B with a covariate on the same set of pupils; and (2) the pupils were more comparable because they were selected from the same percentile ranges. One of the biggest potential problems in any Title I evaluation is that the populations above and below the selection cutoff might have different distributions. Admittedly, while the method by which the experimental sample was

pulled didn't assure equivalent distributions, it did make some effort to control for any differences that might exist between the distributions. The data collected indicated that the efforts had been somewhat successful. Results of the Model C are given in Table 9.

TABLE 9

Predicted, Observed and Differences in NCE Scores for the Experimental Group Subpopulation

	Predicted NCE	Observed NCE	NCE Difference
BSSI*	29.40	49.03	19.63
VAD†	27.85	40.21	12.36
Math$^{\psi}$	39.01	45.59	6.58

* Slope = .43, control group mean = 39.38, difference in selection means = -23.20

† Slope = .78, control group mean = 45.85, difference in selection means = -23.20

ψ Slope = .49, control group mean = 50.38 difference in selection means = -23.20

Difference scores in all cases showed positive program impact, with all differences well above or close to the 7 NCE recommendation for determination of educational significance.

6. A Model A evaluation was performed on the subset of pupils in the treatment group who scored below the cutoff selection score. All of these pupils were part of the experimental study which is mentioned in the modified Model B evaluation previously discussed. It was performed for the reasons cited above. Results showed that the NCE gain from pre to posttesting was 32.66 on the BSSI. As in the Model A evaluation on the entire KED population, these results, being more than 1.5 standard deviations, would indicate tremendous programmatic impact.

Although the use of Model C does not require the computation of the slope of the treatment group, it should be obvious that if the population above the cutoff score has a different distribution of scores than the treatment group, the slope for this group will be different than the slope for the treatment group. To see what effect these differences would have on the predicted or expected post scores of the treatment group, the means on the criterion test for those in the control group, in the experimental study, who scored below the cutoff were compared to the Model C predicted scores for the treatment group. These predicted scores represent the scores that should have been obtained by the treatment group if they had not received services. Since the subset of the control group mentioned above came from the same schools as the treatment group, had scores below the cutoff, did not receive treatment, and since both groups had been randomly selected, the procedure indicated above seemed to be a logical way to test the accuracy of the predictions under Model C. Table 5, presented earlier, gives the slopes for the appropriate subgroups in the experimental study. The slopes for the BSSI and the Math subtest of the CTBS with the selection test are relatively close for both groups (treatment and control), but the slopes for the VAD subtest of the CTBS appeared to be quite different. Table 10 shows the predicted NCE scores for the treatment group if no services had been given, the observed NCE scores of the control group below the cutoff who didn't receive Title I services, and the NCE difference (Observed - Predicted) score for the control group of pupils.

TABLE 10

Predicted, Observed, and Difference NCE Scores
(Control group below cutoff)

	Predicted NCE	Observed NCE	NCE Difference
BSSI	29.40	29.34	- .06
VAD	27.85	34.00	6.15
Math	39.01	39.56	.55

Inspection of Table 10 shows that for the BSSI and the Math subtest of the CTBS, where the slopes were similar, the Model C predicted the mean of the control group quite well. The Model C prediction on the VAD subtest of the CTBS, where the slopes were dissimilar for the two groups with the slope for the control group being higher, under predicted the mean of the no treatment control group. If the slope for the control group had been as much lower than the treatment group's slope as it was higher for VAD, Model C would have overpredicted the posttest mean of the no treatment control group below the cutoff by 2.67 NCE's (i.e., NCE difference would have been -2.67). Therefore, based on the evidence of the experimental study presented in this paper, when the slopes on the criterion variable(s) with the selection variable are similar, Model C seems to adequately predict the mean score without treatment on the criterion variable(s).

DISCUSSION

It should be obvious from the previous sections of this paper that the USOE models produce substantially different results from one model to another. For example, the estimate of the NCE gain on the BSSI for the treatment group of the experimental study is shown in Table 11 for each model used.

TABLE 11

Treatment Effects on the Criterion Variable BSSI, NCE Gain, by Model (Experimental Study)

Model	*Treatment Group NCE Gain*
A	*32.66*
Modified B	*14.77*
Modified B with Covariate	*21.62*
C	*19.63*

The difference between Model A and the modified Model B was 16.89 or eight-tenths (.8) of a standard deviation. In fact, Model A produced the most favorable results of programmatic impact in all cases. The minimum difference was 11.04 NCE's or slightly more than .5 standard deviations.

Aggregating evaluation results across the models, then, is comparable to comparing the weight (not mass) of an object on the surface of the Moon to its weight on the land surface of Mars. The resultant readings would be highly discrepant, and without additional information about where the measurements had been taken, different conclusions about the actual mass of the object would be drawn. The impact of an educational program is like the mass of an object. The mass of the object is fixed and does not vary whether it is on the Earth, Mars, or the Moon, but the weight of the object is highly affected by its location. Similarly, for any given year, the educational impact of a program is fixed, but the different models certainly give different readings (weights) as to the extent of its impact. For example, if a third of a standard deviation (7 NCE's) is used as the standard for determining educational significance, as recommended by the model developers, the effectiveness of the KED program, on the criterion variable VAD subtest of the CTBS, would vary by the model employed. For example, on the 507 pupils of the experimental study the modified Model B shows no positive treatment effect at all (-1.06 NCE, treatment-control difference); the modified B with covariance shows positive (5.81 NCE difference), but not educationally significant results; and Model C shows positive (12.36 NCE difference) and educationally significant results.

Now if these results above had come from three different school districts (or states) which were subsequently added, averaged, and then reported to Congress, Congress would be lead to believe that the program was educationally insignificent, but not necessarily worthless. If the results were not aggregated,

but reported to Congress, then Congress would conclude that one system's program was ineffective, another effective (but not significant), and the third was positive and educationally significant when the only real difference was not the effectiveness of the program but the model used to evaluate the program.

In addition to the noncomparability of the models, the use of Model C showed substantially different results on criterion measure when the slope of line for those pupils above the cutoff was not similar to the slope of the line for the pupils below the cutoff. For example, the Model C evaluation on the experimental group varied on the VAD and Math subtests of the CTBS, from that on the entire KED population. Table 12 shows the results of the Model C evaluation on both groups and the slopes for those pupils above and below the cutoff score.

TABLE 12

Comparison of Model C Results on the VAD & Math Subtests of the CTBS for the Experimental Study & Entire KED Population

	VAD			Math		
		Slope			Slope	
	NCE Gain*	KED Below Cutoff	Non-KED Above Cutoff	NCE Gain*	KED Below Cutoff	Non-Ked Above Cutoff
Experimental Study	12.36	.59	.78	6.58	.41	.49
All KED	7.40	.59	.47	2.10	.54	.34

* Observed NCE - Predicted NCE

As mentioned earlier in this paper when the slope is greater for the no treatment group than it is for the treatment group, Model C will underpredict the observed NCE of the treatment group which will produce an overestimate of the treatment effect (see VAD results in Table 12). When the slope of the no treatment group

is less than it is for the treatment group, Model C will overpredict the observed NCE of the treatment group which will produce an underestimate of programmatic effect (see Math results in Table 12). The data in Table 12 combined with that of Table 10 illustrate the phenomenon described. The Model C estimates of programmatic impact of the KED program are probably best described for the VAD-CTBS subtest when performed on all KED pupils and for the Math-CTBS subtest when the experimental study group is used.

RECOMMENDATIONS

Based upon the findings of this study, the following specific recommendations are made:

1. Results of Model A evaluations must be viewed with utmost caution. Due to the model's tendency to "overestimate" the actual effect of a program, an NCE of "7" should not be viewed as educationally significant as an NCE difference of 7 in either Models B or C.
2. USOE should encourage and permit the implementation of Model B procedures in selected school districts. Since it is unknown what the actual effects of programs are until after the evaluations have been completed, there is no reason to assume the conditions required in a Model B evaluation are in any way depriving anyone of anything that is desirable. In fact, the lack of sustained or systemmatic effects of most Title I programs (at least in St. Louis), would suggest that implementation of Model B may be the most logical and perhaps the only way to determine a program's effectiveness.
3. Before one determines the effectiveness of educational programs by Model C procedures, it should be required to compute and test the equality of slopes on both pupils above the cutoff who receive no treatment and pupils below the cutoff who receive services. From the data in this study, it would

seem, that in most instances where the slopes for the two groups are within one-tenth (.1) of each other, Model C produces fairly accurate results.

4. USOE, state and local educational agencies should not aggregate results across different models. The differences in programmatic impact, measured in terms of NCE's, are highly discrepant from one model to another.

5. It was hoped that the development of the NCE metric would permit comparison of achievement data across and within models. It should be apparent, however, that the adoption by itself of a standard metric will not equate evaluation findings regardless of procedures. Additionally, the NCE metric as presently used is not without problems. In order to determine the value of a particular education program the development of additional evaluation approaches and models must be encouraged.

6. Empirical studies to determine comparability, advantages and disadvantages of the different models must continue to be conducted. Until such information is collected and evaluated, USOE should rethink its present policy of requiring implementation of the RMC models in Title I evaluations.